PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification [6] :		(11) International Publication Number: **WO 95/20291**
H04N 1/32, 1/00	**A1**	(43) International Publication Date: 27 July 1995 (27.07.95)

(21) International Application Number: PCT/GB95/00099

(22) International Filing Date: 19 January 1995 (19.01.95)

(30) Priority Data:
9400971.9   19 January 1994 (19.01.94)   GB

(71) Applicant (for all designated States except US): MOR LTD. [GB/GB]; 1 Hanbury Mews, Mary Street, London N1 7DL (GB).

(72) Inventors; and
(75) Inventors/Applicants (for US only): PAATELMA, Otso [FI/GB]; 1 Hanbury Mews, Mary Street, London N1 7DL (GB). BORLAND, Rod, Hugh [ZA/GB]; 1 Hanbury Mews, Mary Street, London N1 7DL (GB).

(74) Agent: ORIGIN LTD.; 1 Hanbury Mews, Mary Street, London N1 7DL (GB).

(81) Designated States: AU, CA, CN, GB, JP, KR, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHOD OF AND APPARATUS FOR MANIPULATING DIGITAL DATA WORKS

(57) Abstract

A method of hiding copyright related messages within digital data works is taught. The method relies on commonly occurring patterns or sequences of data elements in the work acting as signpost to target data elements which are modified according to certain rules; the presence of the modification indicates that the work is a copyright work; the actual nature of the modification can code for textual information, enabling more detailed information, e.g. the name of the copyright holder, to be hidden within the work.

| 1 |

| a | x |

| x + b |

1

Method of and apparatus for manipulating digital data works

Field of the invention

5

This invention relates to a method of and apparatus for manipulating digital data works, particularly to provide, or embed within. such digital data works, data information descriptive of the work, for example. the source of the work. its ownership and its availability for legal reproduction without infringement of copyright.

10

Description of the Prior Art

Systems based upon digital data are becoming universal and indispensable; digital data
15   passing between computers: digital telecommunications: digital audio: digital cameras; and the convergence of many of these individual components into multi-media, are a selection of the technologies to which this invention relates. The data structures or formats that are used in these different technologies are well documented and will not be described in this specification. It will however be appreciated by the skilled
20   implementer that sophisticated encoding and compression algorithms are commonly used in digital recording and transmission and that these techniques may involve the manipulation of raw digital data in order for that data to carry further information. However. the further information introduced in such known techniques provides data needed to understand or handle. for example. bytes of the raw data properly. The
25   present invention is not directed to manipulating small units of digital data to enable that data to carry information inherent to the proper comprehension of the digital data itself but instead to hiding within a digital data work information that is specifically descriptive of that particular digital data work. The term digital data work defines any sequence of digital data capable of constituting a work in which copyright might subsist
30   or the unauthorised reproduction of which might constitute infringement of copyright subsisting in the work or a larger work of which the work itself forms a part. Hence, for example, the digital data defining the output from scanning a 35mm colour transparency would be a digital data work.

35   There is a pressing need to enable the ownership of a digital data work to be readily apparent from the work itself. This would allow the owner of the copyright in the data to prove more readily its ownership and thereby prevent further unlawful reproduction

2

or negotiate a licence fee for use of the work. Also, it would enable a publisher of digital data works, for example a company that compiles multi-media CD-ROMS, to ensure that it is not unwittingly infringing copyright. Conventionally, this requires laborious and meticulous manual recording of the provenance of third party digital data

5   works that are to be used. This is time consuming, expensive and not wholly reliable.

As a more specific example, a professional photographer uses a digital camera to take a digital photograph. He may wish to place the data file which constitutes the digital data work (the digitised photograph) with a photograph library. When the photograph is

10   subsequently used, the photographer will be paid a copyright licence fee. However, the ease with which the data file can be duplicated without any loss of information means that the further use of the image can be impossible to control. In practice, there is a high probability that the image could be published again without the photographer being alerted to the fact: the photographer losses out on licence fees. The ease with which

15   digital images can be cropped and modified compounds the problem.

Currently, it is possible to include a simple identifier in the header of the data file of a digital data work. The header could typically comprise a simple copyright notice. However, it is very easy to strip out this information, making this approach unreliable.

20   Further, translating a file into a different format also results in this information being lost.

Statement of the invention

25

In accordance with the present invention, a method of manipulating a digital data work, made up of a number of data elements, to include additional data descriptive of that digital data work, comprises the steps of:-

analysing the digital data work for the occurrence of a particular pattern or

30   sequence of data elements in the work, such a pattern or sequence being indicative of the location of a target data element in the work; and

modifying characteristics of the target data element in the work in accordance with a predetermined rule, to generate the additional data, the presence of the additional data not being readily apparent or detectable in the digital data work after manipulation.

35

In this way, the additional data which is descriptive of the digital data work, typically a copyright notice or more detailed historical information broadly pertaining to the

copyright in the digital data work (for example, date of creation, identity of creator etc), can be hidden within the digital data work in a manner that is not readily detectable and therefore does not discernibly alter the digital data work. Being able to hide such data within a work clearly also has attractions to the non-commercial user as well; for example, many cameras today have the facility to imprint a date on an image. This has the disadvantage of defacing the image. With the present invention implemented into a digital camera, the photographer can hide not only date information but other data too into each digital image with no appreciable affect to the image itself. Typically, a particular pattern or sequence of data in the digital data work in effect surrounds a candidate or target data element: the location of the manipulated target data elements is therefore apparent if one knows what particular pattern or sequence of data in the digital data work, or portions of that work, to look for. In a simpler variant, the mere presence of manipulated target data elements constitutes the additional data; i.e. there is no textual information per se hidden within the digital data work. Instead, the presence of target data elements manipulated in accordance with a predetermined rule acts as a indicator that the work is. for example. a copyright work which cannot be reproduced without specific permission from the right holder. This can be likened to giving the digital data work a 'fingerprint'.

In more complex variants. the actual data content of the manipulated target data elements may code for particular textual information pertaining to the copyright status of the work. In this way, a copyright notice that would be hidden to persons, for example, viewing a digitised image, can be detected if one knows the particular pattern or sequence of data which is the signpost to a data modification as defined above. Hence, such a digital data work is both 'fingerprinted' and comprises 'encrypted' text.

The particular pattern or sequence may exist as a sequence in time, for example where the digital data work is an audio work. In addition, it may occur as a pattern of locations that are defined by their spatial relationships to one another. This may arise where the digital data work is an image. The pattern may also exist as a combination of temporal and spatial sequences or patterns. for example where the digital data work is an audio-visual work. The elements may be the discrete or basic units of digital information and the characteristics may be the values of those units of digital information. For example. where the digital data work is an audio work. the characteristics may be the values of the quantised amplitude levels associated with each successive time period. If the digital data work is a digital image. then the elements may be the pixels and the characteristics may be the pixel values. The elements need not

4

from part of a continuous sequence, nor occur on the same row. In some circumstances, it may be preferable to have a complex algorithm to determine the location, in either time or space, of the elements that may form into the required pattern. This may be particularly useful where there is a need to conceal the required pattern in order to prevent the system being cracked.

Generally, the particular pattern or sequence chosen will be selected because it usually occurs many thousands of times in a typical digital data work; so that there is considerable repetition and therefore redundancy in manipulations. For example, in the context of digital imagery. the pattern may be the presence of a given number of pixel values which are each different from their spatially adjacent neighbours along a given row; the value of the subsequent pixel along the row may then be replaced with the value of the next pixel value along the row. In this way, many manipulations would be carried out for a typical digital image, so that even severe distortions or selections from this image will still result in many sets of additional data within the resulting, manipulated digital data work.

Careful selection of the particular pattern is required to resist some compression techniques. For example, it may be preferable for each element to be a set of pixels rather than a single pixel. Then. the characteristics can be the colour temperature of the set of pixels, the average red value of the set, or any other value attributable to the set. The particular pattern can then be one associated with different sets of pixels. For example, it could be based on sets of four pixels: the pattern could be a sequence of five series of such four pixel sets in which the red average value for each four pixel set increases by more than. say. 10 units. from set to set. The next set of four pixels could be the candidate or target set to be manipulated. The actual manipulation could be the adjustment of the red average value of that next set to be equal to that of the last of the four pixels.

Since the additional data associated with the manipulated elements may represent numbers or letters. text messages can be buried within the manipulated digital data work. In this way, a digital data work can carry in parallel two or more sets of data, one of which is concealed from normal usage of the digital data work.

In one embodiment of this aspect. the digital data work is an image and comprises elements each of which defines a pixel. The particular pattern is the presence, firstly, of a given number of elements having pixel values. i.e. characteristics. which are each

5

different from their adjacent neighbours and which determine the position of a target
element and secondly, the presence, surrounding the target element, of n elements, each
having pixel values that are all different from one another, each representing an n-based
number. The predetermined rule is then that if the target element is to code for the
number m, falling within the range $0 - [n-1]$, then it is given a pixel value equal to that
occurring at the surrounding element associated with the n-based number equal to m.

In either aspect, the predetermined rule may include a sub-rule that specifies that no
modification shall occur if the original and modified elements would differ by more
than a defined tolerance. This ensure that the manipulation of the digital data work does
not alter the image beyond tolerance limits, hence ensuring that the manipulation is not
readily discernible.

In a yet further aspect of the present invention, there is provided a method of reading a
digital data work comprising the steps of analysing the digital data work for a particular
pattern or sequence of data and determining if elements of the work have been replaced
in dependence on the occurrence of the particular pattern of data and in accordance with
a predetermined rule.

In addition, there is provided apparatus for manipulating a digital data work to include
additional data, operable to perform the methods described above.

Brief Description of the Drawings

The invention will now be described with reference to the accompanying drawings, in
which:-

Figure 1 depicts in schematic form the layout of pixels that may be coded for in
one aspect of the invention, to be referred to as "fingerprinting";

Figure 2 depicts in schematic form the layout of pixels that may be coded for in
another aspect of the invention, to be referred to as "encrypting".

6

<u>Detailed Description</u>

The detailed description that follows is in respect of an embodiment of the invention that relates to manipulating the digital data of a colour bit-mapped digital image.

5

The technology of colour bit-mapped imagery is well known and will not be described in detail. For the present purposes, it will suffice to observe that a 4-bit image can comprise pixels with 16 different colours, an 8-bit image 256 and a 24-bit image 16,777,216 different colours. On a computer display, colours are represented by the three colours red, green and blue. Any colour can be generated by mixing the relative amounts of each of these colours for any given pixel. The most convincing representation requires a 24-bit image, where each sub-colour would have a one byte (8 bit) value. However, due to computer hardware speed and price constraints, an alternative system called 'palette' has been developed. A 'palette' is typically an array of 256 3-byte RGB values, defined individually for each image. Using the palette system, the matrix of dots, or bitmap. that forms the entire image does not have to be a 24-bit image but instead just an 8-bit image, where the 8-bit pixel values are pointers to the palette of 24-bit colour values.

20   The present embodiment is implemented in an apparatus, referred to as DIP (Digital Image Protection) which [1] alters digital bit-mapped images by selectively modifying the pixels in the image and [2] scans images for previously performed modifications. One purpose of DIP is to hide text information in the image in such a way that the legal owner of the copyright in the image can prove his/her ownership of the image even after severe modifications of the image. This process is referred to as 'encryption'. Additionally, it allows the image to be modified in such a way that third parties can search for non-textual modifications. the presence of which indicates that reproduction of the image would infringe copyright. This can be thought of as giving a 'fingerprint' to the image. Both the 'fingerprint' and 'encrypted' information is camouflaged in the picture so that the human eye cannot spot the difference between a manipulated image and the original image when they are placed beside each other.

The DIP system operates as follows. Firstly, the bit mapped image is read into an internal buffer in a separate analysis module. hence allowing the image to be of any possible format. The DIP system currently produced assumes that the image has a palette look-up table for colours. but with slight modifications (which would be within the expertise of the skilled implementer) it can work with non-palette images. The

module than determines the run-time parameter, and depending on the task, either modifies the image by encrypting or fingerprinting it, or searches the image data for previously planted fingerprints/encrypted characters.

5   Fingerprinting
A pixel will qualify as a target pixel for a fingerprint if it is preceded by a given number of consecutive pixels along a row which are each different to their immediate neighbours. The value of the pixel a pre-set number of pixels after the target pixel is then determined, and in the case of fingerprinting, the target pixel is given this new
10   value as a replacement for its original value. The replacement does not, however, take place if the target pixel needs to be modified by an amount in excess of a given tolerance. The colour difference is checked with all colours, i.e. red, green and blue, individually.

15   When the DIP system searches for fingerprints, it looks for target pixels that satisfy the above relation, i.e. are pixels that are preceded by a given number of different pixels; if the target pixel's value equals that of the next pixel along the row, a 'found' counter is incremented by one. At the end of the search procedure, a percentage figure of 'found' pixels out of all target pixels is calculated. DIP then displays this figure so that an
20   operator can assess whether or not the image had been fingerprinted.

Referring to figure 1, a pixel [x] in the image is to be modified when [a] previous pixels (1....a) have different values compared to each other. In such a case, the pixel's value is a candidate to be changed to a value equal to that of the pixel [b] pixels away
25   from the target pixel, i.e. pixel [x +b].

To secure the integrity of an image an additional Purity Check ["PC"] is performed when a pixel modification is about to take place. The DIP computer checks the colour [RGB] value difference between the original pixel and the pixel that is proposed for its
30   place. If the value is greater than a defined tolerance, the original pixel is not altered.

Fingerprint Verification
An identical algorithm is used to check an image for previously planted fingerprints. The target pixel's [x] value is verified with the possible replacement pixel's [x + b]
35   value. If the values match, it is likely that pixel is fingerprinted. After collecting a large amount of targets the probability of random matching reduces. In a scanned photographic image, the amount of random matches in non-fingerprinted images is

8

usually between 10% and 20%, whereas in a fingerprinted image the percentage is naturally 100% or very close to that.

Encryption

5   The encryption of alpha-numeric characters also starts from a target pixel, i.e. one for which a given number of consecutive preceding pixels are each different to their neighbours. An additional condition is also checked: it is whether or not a particular array of four surrounding pixels all have a different value to each other. If this condition is fulfilled, the candidate pixel will swap its value with one of the four pixels

10   around it, so long as the colour tolerance allows it, i.e. the change is within a defined tolerance. Hence the target pixel will code for a four-based number. These numbers are then fed to a six-digit array. The first three digits are the actual figures, with a value ranging from 0 to 63 [63 being 4^3 -1]. The next number will be a checksum, or parity and the last two are end of character indicators, or stop bits. A figure 32 is added to the

15   resulting figure so the result can be an ASCII character between 32 [space] and 95 [underscore]. This includes all uppercase English letters, numbers and most common punctuation marks. In this way, successive encrypted pixels can define any alpha-numeric character.

20   Very rarely, a candidate or target pixel is found, but the target pixel happens to be the same as one of the four surrounding pixels. DIP will then change the candidate pixel to another value altogether, if the tolerance will allow, so that when decrypting the image DIP will understand that this pixel cannot carry any information.

25   DIP converts the ASCII characters of the message to four based numbers one by one; the checksum, or parity number and the end-of character indicators, or stopbits, are then added to these figures. Messages can include copyright notices, or code numbers that are recorded on a central database against legal owners etc. of the image. In this way, a third party wanting to reproduce the image need only use the DIP system to

30   extract the relevant code number and then communicate it to the central database controller, who in turn will inform the third party of the relevant person to whom royalties need be paid. In this way, the problem of researching who owns an image, which can be very labour intensive. is entirely obviated.

35

Returning to the DIP system. DIP typically fingerprints lines 0. 2, 4, ... and encrypts lines 1, 3, 5 ... of the image. When searching for the fingerprints, it scans all the lines

9

and determines, depending on the quantity of found fingerprints, if the encryption codes should be looked for in the odd or even lines.

5   The encrypted characters, if found, are printed on the screen of the DIP. On termination, the program analyses the probability for the image having been fingerprinted earlier by DIP, the analysis being based on the percentage of 'found' pixels, and prints a verbal message as a result.

More generally, and referring to figure 2, the encryption concept uses fingerprinting
10   with Additional Condition Verifying (ACV) to store a fragment of an [n] based number at that location. An ACV condition is true when [n] pixels near the location are all different, hence enabling the pixel at the location to be uniquely modified to reflect one of the pixels around/near it. An example of such conditions follows. A pixel [x] in the image is to be modified when [a] pixels (1...a) have different values compared to each
15   other and [n] surrounding pixels (P1.....Pn) have different values to each other. The new value of the pixel [x] will be one of the pixels P1....Pn, hence identifying an n-based number (0.....n-1). If the number to be stored at that location is 0, then target pixel [x] is given the value of pixel P1. If the number to be stored is 1, then it is changed to the value of pixel P2, and so on. This change will obey the same PC
20   limitations as described earlier.

By combining these n-based numbers sequentially and organising them in an array, larger figures can be constructed. When writing the numbers into an image, it is useful to construct the array of numbers in such a way that it has a built-in checksum and
25   figure-stop indicator trailing each number. This enables DIP to decode information from an image that has been altered.

Decoding of Encrypted Messages

30   An identical algorithm is used to decode an encoded image. When a candidate pixel [x] is found with [n] pixels around it all different, its value is verified from the nearby pixels. The ASCII character associated with the value of the number is found by reversing the process used when encoding.

35

10

Claims

5

1.   A method of manipulating a digital data work, made up of a number of data
elements, to include additional data descriptive of that digital data work, comprising the
steps of:-

10   analysing the digital data work for the occurrence of a particular pattern or
sequence of data elements in the work, such a pattern or sequence being indicative of
the location of a target data element in the work; and

modifying characteristics of the target data element in the work in accordance
with a predetermined rule, to generate the additional data, the presence of the additional

15   data not being readily apparent or detectable in the digital data work after manipulation.

2.   The method of Claim 1 where the particular pattern or sequence of data enables
the target data elements to be unambiguously identifiable.

20   3.   The method of either Claim 1 or Claim 2 wherein the digital data work is an
image and each data element is a pixel or a set of pixels.

4.   The method of Claim 3 wherein: the particular pattern is the presence of a given
number of data elements having characteristics which are each different from their

25   adjacent neighbours by more than a predetermined amount; the target data element is in
a predetermined location with respect to such data elements and the predetermined rule
is that the characteristic of the target data element is replaced by the characteristic of a
predetermined data element.

30   5.   The method of Claim 1, 2 or 3 wherein the step of modifying a characteristic of
a target data element leads to the presence of additional data that codes for a particular
number or letter.

6.   The method of Claim 5 wherein the value of the modified target data element

35   has a predetermined relationship to another data element and the location of that other
data element with respect to the target data element is indicative of a number or letter.

7.   The method of Claim 6 wherein the value of the modified target data element equals that of the other data element and the number of elements separating the target data element from the other element codes for a particular number or letter.

5   8.   The method of Claim 6 wherein the digital data work is an image and comprises elements each of which is a pixel or group of pixels and the particular pattern is the presence, firstly, of a given number of elements having values of a variable which are each different from their adjacent neighbours and secondly, the presence, surrounding the target element, of n elements, each having values of that variable that are all

10   different from one another, each representing an n-based number, and wherein the predetermined rule is that if the target element is to code for the number m, falling within the range 0 - [n-1], then it is given a value for its variable equal to that occurring at the surrounding element associated with the n-based number equal to m.

15   9.   The method of any preceding Claim wherein the predetermined rule includes a sub-rule that specifies that no modification shall occur if the characteristics of the original and the replaced elements would differ by more than a predetermined amount.

10.   A method of reading a digital data work to determine if the work has been
20   manipulated in accordance with any preceding Claim 1-9.

11.   Apparatus for manipulating a digital data work to include additional data, characterised in being able to perform the method of Claims 1-9.

25   12.   Apparatus for reading a digital data work, operable to perform the method of Claim 10 and to display a result indicative of whether the digital data work has been manipulated.
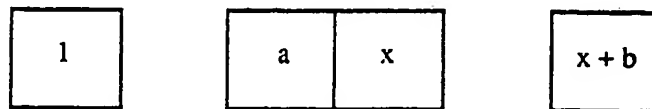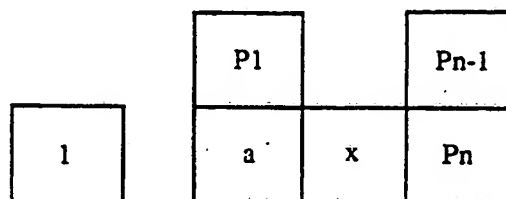
30

35

1

a	x

x + b

Figure 1

P1		Pn-1

1

a	x	Pn

Figure 2

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6   H04N1/32   H04N1/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6   H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP,A,0 551 016 (CANON KK) 14 July 1993 see column 4, line 44 - column 7, line 52 ---	1
A	ELECTRONICS AND COMMUNICATIONS IN JAPAN, vol. 73,no. 5, 31 May 1990 NEW YORK (US), pages 22-33, XP 000159282 KOMATSU N. ET AL 'A PROPOSAL ON DIGITAL WATERMARK IN DOCUMENT IMAGE COMMUNICATION AND ITS APPLICATION TO REALIZING A SIGNATURE' see the whole document ---	1
A	EP,A,0 483 936 (IBM) 6 May 1992 see abstract ---	1
A	EP,A,0 532 381 (GEMPLUS CARD INT) 17 March 1993 see abstract ------	1

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 May 1995	2 4. 05. 95

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Hazel, J

Form PCT/ISA/210 (second sheet) (July 1992)

Inter   nal Application No
PCT/GB 95/00099

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
EP-A-0551016	14-07-93	CA-A-	2086472	07-07-93
		JP-A-	5301380	16-11-93
EP-A-0483936	06-05-92	US-A-	5227893	13-07-93
		JP-A-	4333994	20-11-92
EP-A-0532381	17-03-93	FR-A-	2681490	19-03-93
		DE-D-	69200087	05-05-94
		DE-T-	69200087	01-09-94
		ES-T-	2056698	01-10-94
		JP-A-	5244441	21-09-93